BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e139153
doi: 10.3897/biss.8.139153

OPEN ACCESS

Conference Abstract

The Dream of Mainstream in TDWG, GBIF and SPNHC

Arturo H. Ariño ‡

‡ BIOMA Biodiversity and Environment Institute, University of Navarra, Pamplona, Spain

Corresponding author: Arturo H. Ariño (artarip@unav.es)

Received: 12 Oct 2024 | Published: 14 Oct 2024

Citation: Ariño AH (2024) The Dream of Mainstream in TDWG, GBIF and SPNHC. Biodiversity Information Science and Standards 8: e139153. https://doi.org/10.3897/biss.8.139153

Abstract

The Biodiversity Information Standards (TDWG), the Society for the Preservation of Natural History Collections (SPNHC) and the Global Biodiversity Information Facility (GBIF) have been providing nuts and bolts to biodiversity research for decades. Standards service the research community while requiring significant research themselves. Annual conferences gather a rather persistent group of practitioners who showcase the workings and products of the community (Ung and Kampmeier 2022). Is this effort visible to, and recognisable by, potential stakeholders?

One way to put our groups in perspective is to look at the mainstream scientific production using theirproducts. GBIF has long been keeping track of research using the records it mediates, as well as monitoring the literature citing GBIF itself. TDWG, on the other hand, while also keeping record of its history, does not have a similar mechanism —perhaps because TDWG's (and SPNHC's) usual outlets often may not be on a par with the preferred scientific venues: standard, peer-reviewed, indexed papers. Amazingly enough, a huge amount of work seems to be known only through conference abstracts, presentations and posters, which even though actually peer-reviewed, do not truly conform to the "gold standard" of scientific publications, and are often contemplated by many actors such as evaluation agencies as merely grey literature representing "fringe" research.

I have attempted to measure how much "mainstreamliness" TDWG, SPNHC and GBIF carry, by looking at how frequently their outputs show up in indexed research along with

© Ariño A. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

their citation patterns, and comparing those of other examples both related or unrelated to the three organizations' remits.

METHODS

In 2024, I queried Web of Science (WoS), Scopus (SC), and Google Scholar (GS) repositories for papers according to each platform's capabilities, separately targeting (whenever possible) titles, abstracts, keywords, main texts, and references cited (Table 1).

Table 1.

Search strategies and limits. All searches were done separately for the entire corpus and for recent production (2000 onwards). Limits are per query.

	Web of Science (WoS)	**Scopus (SC)**	**Google Scholar (GS)**
Searchable fields	Title, keys, abstract	Title, keys, abstract, literature, conference	Title, all text; abstracts (current year only)
Hit counts	Exact	Exact	Estimate
Exportable records limit	All	20,000	About 800
Citation counts	Per hit	All, per hit	Per hit
Citation report limit	10,000	10,000	From mined records only

Queries were crafted to find output from four groups: one focal (TDWG, GBIF, SPNHC, and the Darwin Core standard, DwC), and three containing examples of related activity or concepts; specific, biodiversity-related research; and unrelated ("outgroup") general research (Table 2). I kept a tally of hits and other data such as citation numbers, and downloaded (or mined in the case of GS) either the full list of references or a sample, depending on the platform's limitation.

Exported references were combined in a database, filtered, and quality-checked. Indexed citation levels were obtained from either complete sets, or 10,000-record samples. The mainstream share was calculated as the quotient between WoS hits and GS hits per query.

I defined the relative balance, or leverage, of the community's uptake as:

$$\text{leverage} = (SCc - SChl) / \max[SCcm\ SChl]$$

where *SCc* is the number of citations reported by Scopus for the indexed hits, and *SChl* is the number of records found by querying the indexed references' literature lists (which contain both indexed and non-indexed literature). The index is positive when fringe literature is cited preferentially by fringe literature, negative when fringe literature is disproportionately cited by indexed literature, and zero when there is no uptake selectivity.

The overlooked (i.e., used but not properly cited) production was estimated from the ratio of papers found by querying titles, keywords and abstracts, and papers found by querying literature citations. Low numbers mean that most indexed papers using certain information get it mostly from other indexed papers and there is little uptake from non-mainstream sources.

Table 2.

Query constructs. Syntax given as examples—specific rules applied to each platform.

Group	Concept	Query examples
Focal	TDWG	Taxonomic databases working group OR TDWG
	SPNHC	Society for * preservation * natural history collections OR SPNHC OR...
	GBIF	GBIF, global biodiversity information facility
	DwC	darwincore OR darwin core
Focal-related	Standards	biodiversity | taxonom* standar*
	BDI	biodiversity informatics
	Databases	biodiversity | taxonom* database*
Biodiversity research	SDM	species distribution model*
	Broad terms	biodiversity, taxonomy, ecology, bioinformatics
	Taxon examples	Sylviidae, Polychaet*, Fagus
Out terms	Biomedical	Clostridium
	Technical	Artificial Intelligence

RESULTS

Data were available for about 9 million (WoS), 16 million (SC), and 23 million (GS) records, of which about 100,000 were used as the analytical sample.

GS revealed a flow of focal scholarly products growing at different rates. Detected GBIF output grew exponentially, doubling every 2.9 years, while TDWG and SPNHC have remained approximately constant over the last decade at about 375 and 76 documents per year, respectively (Fig. 1). However, the fraction of indexed output is tiny: 0.5% for TDWG since 2000 (0.8% overall), 0.8%–1% for SPNHC, and 3.9%–3.4% for DwC. GBIF has a higher proportion at 4.9% (recent) and 8.2% (overall). Related products have a recent 5.6% indexed share (6.7% overall), but they are all in stark contrast to other research areas: 62.4% (61.2%) for general concepts, 34.0% (61.2%) for taxonomical terms, and 38.8% (31.6%) for the selected outgroups (Fig. 2).

The leverage showed a marked difference between focal and related areas, and general areas. TDWG, SPNHC and taxonomical databases had strong indexed leverage: their documents were overly cited in indexed literature. GBIF and standards citations were biased otherwise, being preferentially cited in unindexed literature but less so than all of the comparison terms (except Bioinformatics, neutral) (Fig. 3).

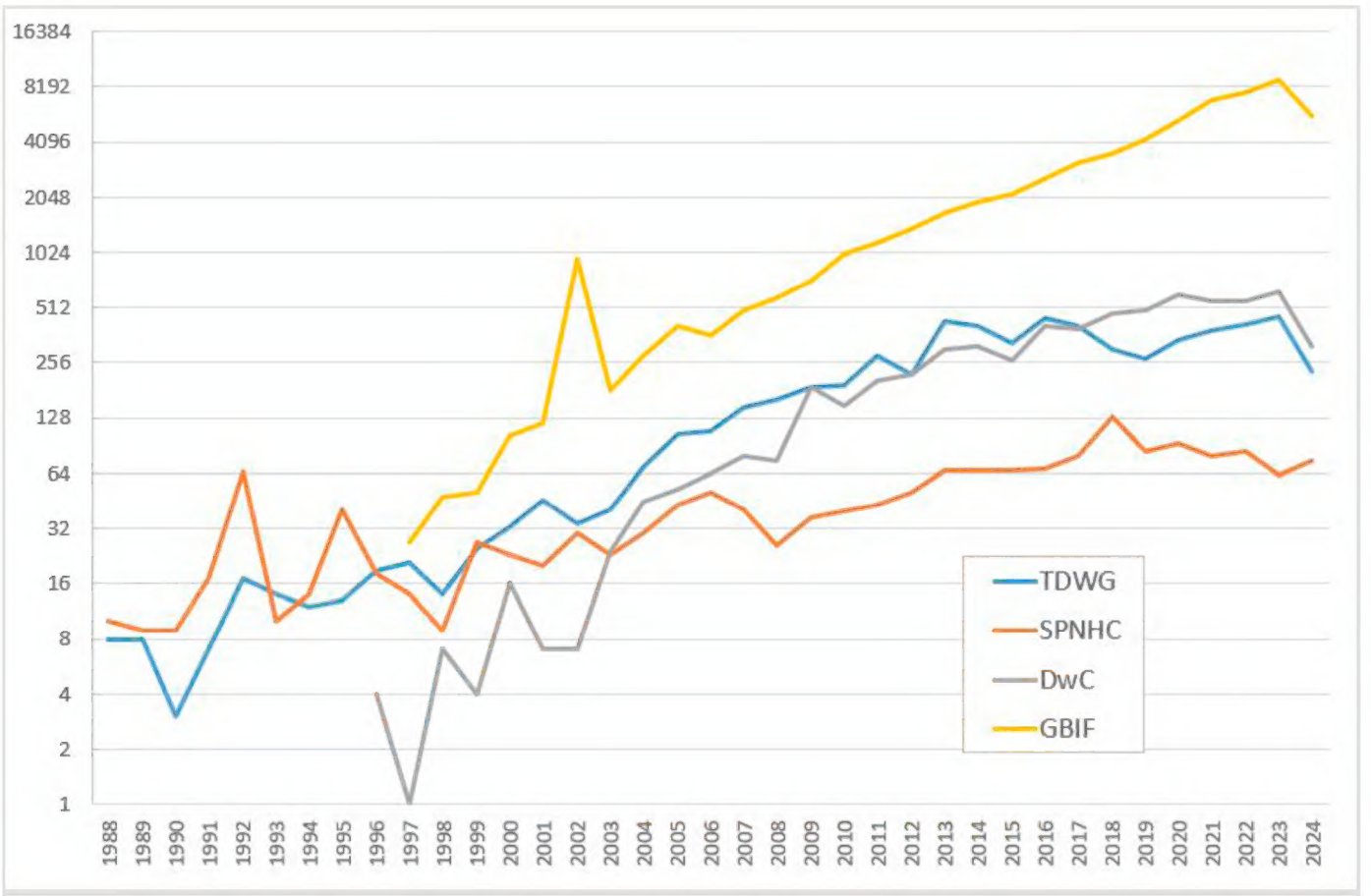

Figure 1.

Documents per year, log(2) scale.

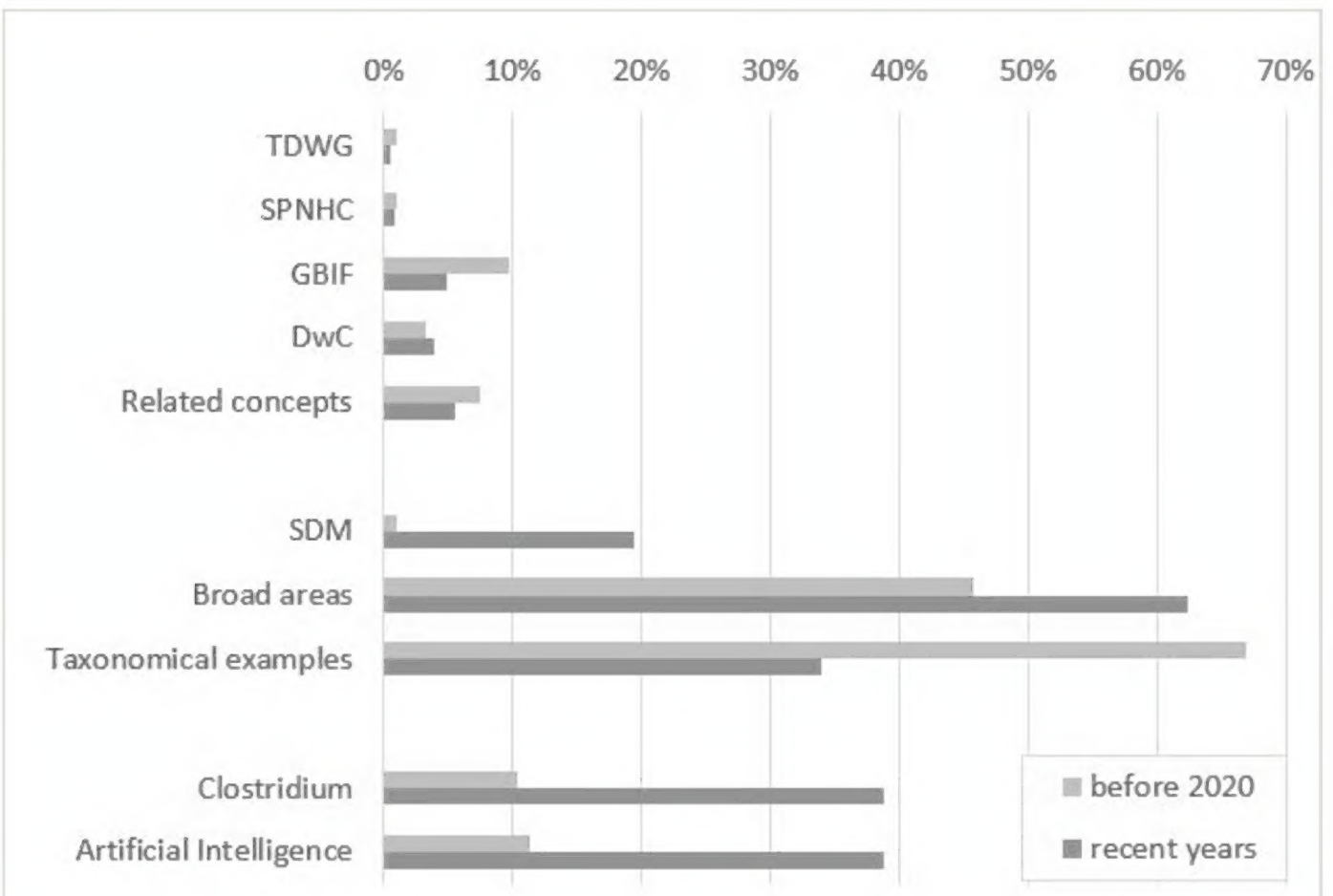

Figure 2.

Estimated share of Web of Science-indexed documents.

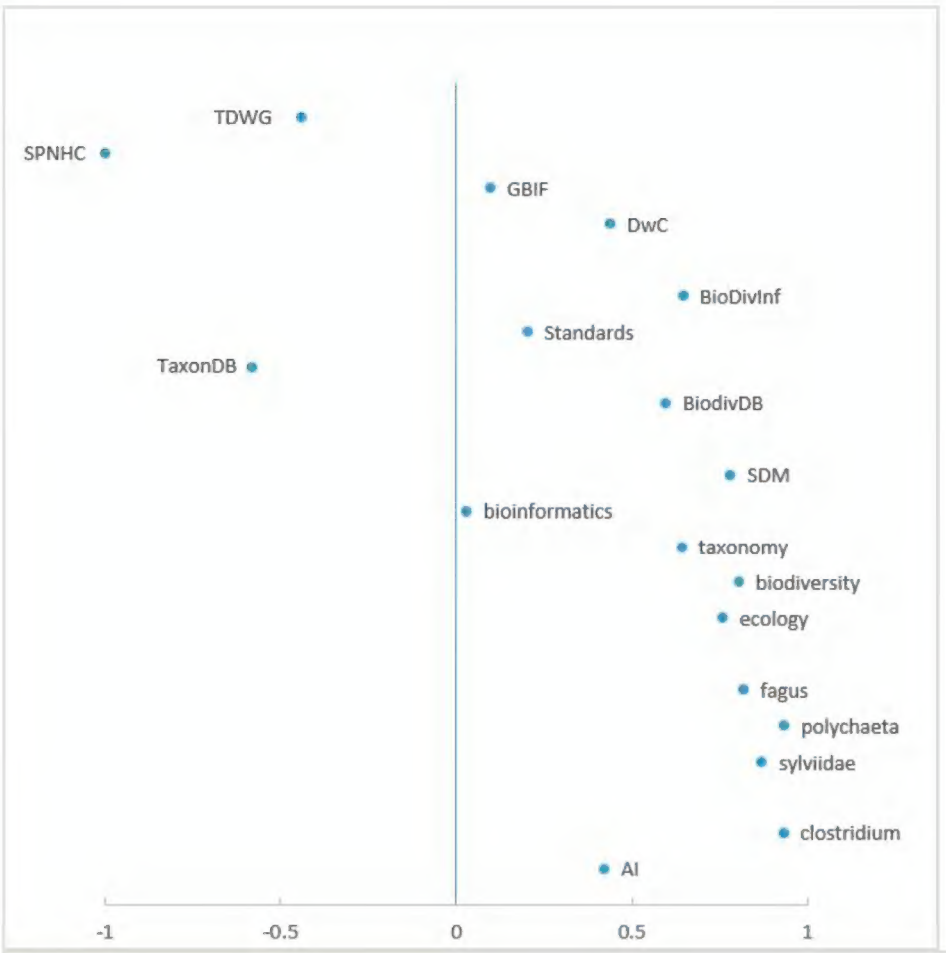

Figure 3.

Citation leverage. Negative: cited papers appear proportionally more in indexed literature. Positive: appearing more in unindexed literature.

While almost one-fourth of Scopus-indexed TDWG literature came from conference papers, these tended to be cited in SC-indexed articles rather than other conference papers in a 1:4 ratio. GBIF or DwC showed a more similar distribution of main types (articles, book chapters, conference papers, reviews) between published documents and cited documents.

A reasonable conclusion is that despite the low proportion of indexed publications by TDWG, SPNHC or GBIF, their products are indeed uptaken by indexed publications, and comparatively much more so than in other areas. Thus, the scientific or technical production by those organizations does have a recognizable impact and should safely be considered de facto part of the mainstream scientific endeavor.

Keywords

TDWG, SPNHC, citation analysis, scientific impact

Presenting author

Arturo H. Ariño

Presented at

SPNHC-TDWG 2024

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Ung V, Kampmeier G (2022) The Sociological History of Transforming TDWG to Biodiversity Information Standards (TDWG). Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.93866